



AFRL-RX-WP-TP-2012-0355

STATISTICAL ANALYSIS OF HIT/MISS DATA (PREPRINT)

Jeremy S. Knopp

**Materials State Awareness & Supportability Branch
Structural Materials Division**

Li Zeng

University of Texas at Arlington

July 2012

Interim

Approved for public release; distribution unlimited.

See additional restrictions described on inside pages

STINFO COPY

**AIR FORCE RESEARCH LABORATORY
MATERIALS AND MANUFACTURING DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7750
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

REPORT DOCUMENTATION PAGE					<i>Form Approved</i> OMB No. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.						
1. REPORT DATE (DD-MM-YY) July 2012		2. REPORT TYPE Technical Paper		3. DATES COVERED (From - To) 1 June 2012 – 1 July 2012		
4. TITLE AND SUBTITLE STATISTICAL ANALYSIS OF HIT/MISS DATA (PREPRINT)				5a. CONTRACT NUMBER In-house		
				5b. GRANT NUMBER		
				5c. PROGRAM ELEMENT NUMBER 62102F		
6. AUTHOR(S) Jeremy S. Knopp (RXCA) Li Zeng (University of Texas at Arlington)				5d. PROJECT NUMBER 4349		
				5e. TASK NUMBER 40		
				5f. WORK UNIT NUMBER LP110200		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Materials State Awareness & Supportability Branch Structural Materials Division Air Force Research Laboratory, Materials and Manufacturing Directorate Wright-Patterson Air Force Base, OH 45433-7750 Air Force Materiel Command, United States Air Force				8. PERFORMING ORGANIZATION REPORT NUMBER AFRL-RX-WP-TP-2012-0355		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory Materials and Manufacturing Directorate Wright-Patterson Air Force Base, OH 45433-7750 Air Force Materiel Command United States Air Force				10. SPONSORING/MONITORING AGENCY ACRONYM(S) AFRL/RXCA		
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S) AFRL-RX-WP-TP-2012-0355		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited. Preprint to be submitted to Materials Evaluation.						
13. SUPPLEMENTARY NOTES The U.S. Government is joint author of this work and has the right to use, modify, reproduce, release, perform, display, or disclose the work. PA Case Number and clearance date: 88ABW-2012-2090, 5 April 2012. This document contains color.						
14. ABSTRACT There has been renewed interest in the analysis of Hit/Miss or Bernoulli data in the context of Nondestructive Evaluation (NDE). Many contributions have been made with a focus on confidence bound estimation on this type of data. Some concern regarding the proper calculation and use of a90 and a90/95 estimates has been raised. In particular, the behavior of a Probability of Detection (POD) curve for large flaw sizes above stated a90 estimates has been of concern. This paper will give a brief historical overview of the analysis of hit/miss data, and propose a remedy regarding recent concerns. The end result will be a procedure that helps avoid pitfalls in the analysis of hit/miss data.						
15. SUBJECT TERMS probability of detection, hit/miss, bernoulli data, bayesian methods						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT: SAR	NUMBER OF PAGES 16	19a. NAME OF RESPONSIBLE PERSON (Monitor) Mark Blodgett 19b. TELEPHONE NUMBER (Include Area Code) N/A	
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified				

Statistical Analysis of Hit/Miss Data

Jeremy S. Knopp¹ and Li Zeng²

¹Air Force Research Laboratory (AFRL/RXLP),
Wright-Patterson AFB, OH 45433, USA

²University of Texas at Arlington,
Arlington, TX, 76019, USA

ABSTRACT

There has been renewed interest in the analysis of Hit/Miss or Bernoulli data in the context of Nondestructive Evaluation (NDE). Many contributions have been made with a focus on confidence bound estimation on this type of data. Some concern regarding the proper calculation and use of a_{90} and $a_{90/95}$ estimates has been raised. In particular, the behavior of a Probability of Detection (POD) curve for large flaw sizes above stated a_{90} estimates has been of concern. This paper will give a brief historical overview of the analysis of hit/miss data, and propose a remedy regarding recent concerns. The end result will be a procedure that helps avoid pitfalls in the analysis of hit/miss data.

KEYWORDS

Probability of Detection, hit/miss, Bernoulli data, Bayesian methods

INTRODUCTION

Studies of the reliability of Nondestructive Inspections (NDI) began with the United States Air Force (USAF) in the 1960's (Packman et al., 1968) and continued during the 1970's with major efforts conducted by the National Aeronautics and Space Administration (NASA) (Rummel et al., 1974) and the USAF (Lewis et al., 1978). Initial efforts to explain the data from these inspection studies used

binomial statistics (Rummel, 1982). In the 1980's, Berens and Hovey developed parametric methods to analyze both hit/miss and signal response data (Berens et al, 1981) and (Berens, 1989). This provided the foundation for guidance for POD studies that has been adopted by the Department of Defense (DoD), and has been published as a 2nd edition of a handbook (MIL-HDBK-1823A, 2009). Other agencies and industries have also made use of this guidance (Gandossi et al., 2010) and (Drury et al., 2006). It should also be noted that efforts in medical statistics are quite similar to the methods currently used for NDE (Collett, 2002). In medical statistics, the term effective dose is equivalent to a_{50} , and lethal dose is equivalent to a_{90} .

CURRENT METHODS OF ANALYSIS FOR HIT/MISS DATA

The current accepted methods used to analyze hit/miss data are similar to the ones published in an overview paper on the topic (Berens, 1989). In that work, the confidence bounds were global, that is they were calculated to apply to the entire POD curve. Later it was decided that this approach was overly conservative, and it was sufficient to apply confidence bounds for each flaw size locally (Berens, 2000). The likelihood ratio method was also introduced which provided more accurate confidence bounds for hit/miss data (Annis et al., 2007) and (Harding et al., 2003). These two developments were incorporated in the second edition of MIL-HDBK-1823. Recently, concerns have been raised about the current approach, and binomial methods have once again been proposed to mitigate concerns about the behavior of POD for flaw sizes greater than the established a_{90} estimates generated by parametric approaches (Generazio, 2011). Further work on nonparametric methods has been done by Spencer with the assumption that the POD is a continuous non-decreasing function of flaw size (Spencer, 2011).

In the authors' opinion, there is a valid concern raised in (Generazio, 2011), and that is that one could analyze hit/miss data that doesn't meet the requirements of MIL-HDBK-1823A and still generate an estimate for $a_{90/95}$ that could be misused. Another concern in (Generazio, 2011) is the assumption of

a monotonically increasing POD function. Proper 3-point calibration (Rummel, 2005) should mitigate, at least in part, concerns about monotonicity.

A REMEDY FOR CONCERNS ABOUT HIT/MISS ANALYSIS

It is indeed possible to generate estimates of $a_{90/95}$ for hit/miss data that can be misleading. In this paper, we propose a Bayesian approach that involves 3 and 4 parameter models that will provide a remedy for this problem. Generally, specific statistical expertise should be consulted to avoid errors in POD studies.

First, the idea of using more than 2 parameters for the POD model has been proposed in the past (Moore et al., 2001). Additional parameters are added for 2 reasons: 1) the evidence doesn't support the POD curve converging to one for large flaw sizes, and 2) better accounting of false call rates such that the POD curve doesn't converge to 0 for small flaw sizes. The difficulty with conventional methods is in the estimation of the parameters. The Bayesian methods proposed in this work facilitate easy estimation of the model parameters, a simple way to compute confidence bounds, and a systematic method for model selection based on the Bayes factor.

A detailed description of Bayesian methods is beyond the scope of this paper, but an overview of how it can be applied in an NDE context is given in (Thompson, 2010), and all the mathematical details can be found in (Christensen, 2010). A very simple statement of a problem posed in Bayesian terms is shown in Figure 1.

The posterior is a probability density function that literally reads as the probability of model parameters θ , given the data. The likelihood is also a probability density function that is read as the probability of the data given assumptions about the model parameters. The prior can be noninformative as will be the case for this work, or it can represent expert opinion or information from previous experiments. The normalizing constant or marginal likelihood in the denominator is actually integration

over the numerator. Most people are familiar with Bayes rule, but there is a computational challenge in calculating the normalizing constant or evidence as it is commonly called in the literature. Calculating the evidence requires a high dimensional integration that is typically performed via sampling methods such as Markov Chain Monte Carlo (MCMC).

COMPARISON OF BAYESIAN APPROACH WITH STANDARD METHODS

Before applying 3 and 4 parameter models to difficult data sets in (Generazio, 2011), Bayesian methods are illustrated on data from Berens seminal paper on the topic (Berens, 1989). For hit/miss analysis the data is assumed to follow a Bernoulli distribution. Either logit or probit models can be used.

$$y_i \sim \text{Bernoulli}(p_i) \quad (1)$$

$$p_i = \frac{\exp(b_0 + b_1 \ln(a_i))}{1 + \exp(b_0 + b_1 \ln(a_i))} \quad (\text{logit}) \quad (2)$$

$$p_i = \Phi(b_0 + b_1 \ln(a_i)) \quad (\text{probit}) \quad (3)$$

The traditional methods for estimating parameters b_0 and b_1 are already established, and the Bayesian methods will be compared with standard cases. Once again, noninformative priors will be used in this process, and MCMC sampling will be used to extract information about the parameters. The data provided in (Berens, 1989) and displayed in Figure 7 of that work is used as a benchmark. Figure 2 displays the POD curves which appear identical to the ones using the traditional analysis methods. An alternative form of the logit model is given in equation 4, and the parameters μ and σ are related to the parameters in equation 2, by equations 5 and 6. A comparison summary of the results is shown in Table 1.

$$p_i = \left\{ 1 + \exp \left[-\frac{\pi}{\sqrt{3}} \left(\frac{\ln a - \mu}{\sigma} \right) \right] \right\}^{-1} \quad (4)$$

$$\mu = \frac{-b_0}{b_1} \quad (5)$$

$$\sigma = \frac{\pi}{b_1 \sqrt{3}} \quad (6)$$

	Inspector			
Parameter	A	B	C	Composite
$\hat{\mu}$	0.96/0.96	1.11/1.11	0.82/0.82	0.96/0.96
σ	0.59/0.59	1.04/1.05	0.87/0.89	0.88/0.85

Table 1. Original analysis/Bayesian methods

There are no substantial differences in the results for the 2 parameter models. The reason this Bayesian approach is introduced is because the inference needed for 3 and 4 parameter models is quite difficult using traditional methods. The sampling methods in the Bayesian approach are just as easy to implement for the 3 and 4 parameter models as they are for the 2-parameter model.

ADDITIONAL PARAMETERS IN HIT/MISS MODELS

The motivation for adding additional parameters to the model is to better represent false call rates for small flaw sizes, and to better represent the tail behavior of the POD curve for large flaw sizes. Equations 7 and 8 show the form of a 3- parameter model with a lower asymptote for the logit and probit models respectively. Figure 3a depicts this and the value of α can be thought of as a measure of the false call rate. Equations 9 and 10 show the form of a 3-parameter model with an upper asymptote. Figure 3b depicts this 3-parameter model. The β term is a measure of the probability of missing flaws as the flaw size goes to infinity. The 4-parameter model has both terms in it and is described in equations 11 and 12, and depicted in Figure 3c. Proper estimation of the upper asymptote is very important for addressing pitfalls in POD analysis.

$$p_i = \alpha + (1 - \alpha) \cdot \frac{\exp(b_0 + b_1 \log(a_i))}{1 + \exp(b_0 + b_1 \log(a_i))} \quad (\text{logit}) \quad (7)$$

$$p_i = \alpha + (1 - \alpha) \cdot \Phi(b_0 + b_1 \log(a_i)) \quad (\text{probit}) \quad (8)$$

$$p_i = \beta \cdot \frac{\exp(b_0 + b_1 \log(a_i))}{1 + \exp(b_0 + b_1 \log(a_i))} \quad (\text{logit}) \quad (9)$$

$$p_i = \beta \cdot \Phi(b_0 + b_1 \log(a_i)) \quad (\text{probit}) \quad (10)$$

$$p_i = \alpha + (\beta - \alpha) \cdot \frac{\exp(b_0 + b_1 \log(a_i))}{1 + \exp(b_0 + b_1 \log(a_i))} \quad (\text{logit}) \quad (11)$$

$$p_i = \alpha + (\beta - \alpha) \cdot \Phi(b_0 + b_1 \log(a_i)) \quad (\text{probit}) \quad (12)$$

If β is smaller than 0.9, then that means a_{90} and $a_{90/95}$ simply do not exist, even if the 2 parameter model provides estimates. So there is a choice between 4 models: 1) 2-parameter model, 2) 3-parameter model with lower asymptote parameter α , 3) 3-parameter model with upper asymptote parameter β , and 4) 4-parameter model. Each of these can use either the logit or probit link function, so that makes a total of 8 models. The determination of which model is appropriate for a given data set is determined by looking at the ratio of marginal likelihoods a.k.a. the Bayes factor.

AVOIDING PITFALLS IN HIT/MISS ANALYSIS

The combination of the 3 and 4 parameter models with Bayesian methods are powerful tools that can be used to avoid pitfalls in the analysis of hit/miss data. The process is simply to calculate the marginal likelihood for each of the 8 possible models and to determine which model is most appropriate. If the 3 or 4 parameter model is selected, it is very possible that a_{90} or $a_{90/95}$ might not exist even though inference with the 2-parameter model provides estimate. Let's look at an example that appeared in (Generazio, 2011). There was a particular set of data that the author identified as a "Case 2 data set D8001(3)L." It was reported that the "Logit-ML" $a_{90/95}$ value was 12.95 mm. The latest version of MIL-HDBK-1823 was not cited in the paper, so this analysis could have been done with older software with a known deficiency, which was corrected by implementing the likelihood ratio method for confidence bound calculation (Annis et al., 2007). The value obtained for $a_{90/95}$ using the likelihood ratio method is 22 mm and many warnings are displayed throughout the calculation (Annis, 2012).

Since there are many misses for large flaw sizes in this data set, it most likely doesn't meet the requirements to be analyzed according to the procedures set forth in (MIL-HDBK-1823A, 2009). For this case, the data is analyzed using 8 potential models. The 2-parameter, 3-parameter with upper bound, 3-parameter with lower bound, and the 4-parameter model will all be examined for both the logit and probit models. The marginal likelihoods for each are listed in Table 1.

Model type	logit	probit	Bayes factor logit/probit
2-parameter	1.1806e-87	2.7022e-88	4.3690
3-parameter lower bound	2.9183e-89	4.2119e-89	0.6929
3-parameter upper bound	6.1396e-83	8.28e-83	0.7415
4-parameter	9.2848e-83	4.0047e-81	0.0232

Table 1. Marginal likelihoods and Bayes factors for each possible model

The ratio of evidence or marginal likelihoods for competing models is used to select the best model. This ratio is known as the Bayes factor. The marginal likelihood for an individual model doesn't provide useful information, but the relative marginal likelihoods do. Note that all the models with upper asymptotes have far higher marginal likelihoods than models without, so all evidence suggests that an upper bound is necessary for this data set. Also note, that the Bayes factor is only greater than 1 for the 2-parameter case. Since this is the ratio logit/probit, this implies that the probit model better fits the data for all models except the 2-parameter model. The 4-parameter probit model has the highest marginal likelihood compared to all other possibilities, so it is the best choice out of all the models. It turns out

that a_{90} and $a_{90/95}$ do not exist for this data set. The lower asymptote is 0.2657, and the upper asymptote is 0.8574. This basically indicates that the false call rate is quite high, and that the POD on average is 0.8574 as flaw size goes to infinity. Figure 4 displays the 4-parameter probit model which is most appropriate for this data set.

SUMMARY AND CONCLUSIONS

The proper analysis of hit/miss inspection data is still very important since this type of data is still widely collected for POD studies. A brief historical survey of the analysis of this type of data in the context of POD was given and accompanied by recent advances. Some concerns related to the statistical analysis and design of POD studies were listed and addressed. In particular, the potential to determine estimates of a_{90} or $a_{90/95}$ with conventional 2-parameter models when the data do not support their existence was examined. An approach using more parameters in the logit or probit model to adequately describe the tail behavior of POD curves was presented. Bayesian methods can be used to determine parameter estimates for these more complicated models. Finally the Bayes factor is used to determine which model is most suitable, and should mitigate concerns about improperly estimating a_{90} and $a_{90/95}$.

ACKNOWLEDGEMENTS

The authors wish to thank Eric Lindgren for helpful comments that provided useful direction.

REFERENCES

Annis, C., and J.S. Knopp, "Comparing the Effectiveness of $a_{90/95}$ Calculations," *Rev. Prog. Quant. Nondestr. Eval*, Vol 26, 2007, pp. 1767-1774.

Annis, C., “Statistical Best-Practices for Building Probability of Detection (POD) models” R package mh1823, version 2.5.4.1, <http://statisticalengineering.com/mh1823>

Berens, A.P. and P.W. Hovey , “Evaluation of NDE Reliability Characterization,” Final Report AFWAL-TR-81-4160, University of Dayton Research Institute, 1981.

Berens, A.P., NDE Reliability Data Analysis, American Society for Metals Handbook Nondestructive Evaluation and Quality Control, Vol 17, pp. 689-701, ASM International, 1989.

Berens, A.P., “Probability of Detection (POD) Analysis for the Advanced Retirement for Cause (RFC)/ Engine Structural Integrity Program (ENSIP) Nondestructive Evaluation (NDE) System Development,” Technical Report AFRL-ML-WP-TR-2001-4010, Air Force Research Laboratory, 2000.

Christensen, R., W. Johnson, and A. Branscum, *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*,” CRC Press, 2010.

Collett, D., *Modelling Binary Data*, Chapman and Hall, 2002.

Drury, Ghylin, and Holness, Error Analysis and Threat Magnitude for Carry-on Bag Inspection, *Proceedings of the Human Factors and Ergonomic Society 50th Annual Meeting*, Vol 50, 2006, pp. 1189-1193.

Gandossi, L. and K. Simola, “Derivation and Use of Probability of Detection Curves in the Nuclear Industry,” *Insight* Vol 52, 2010, pp. 657-663.

Generazio, E.R. “Validating Design of Experiments for Determining Probability of Detection Capability for Fracture Critical Applications”, *Materials Evaluation*, Vol 69, 2011, pp. 1399-1407.

Harding, C.A., and G.R. Hugo, “Experimental Determination of the Probability of Detection Hit/Miss Data for Small Data Sets”, *Rev. Prog. Quant. Nondestr. Eval*, Vol 22, 2003, pp. 1823-1844.

Lewis, W.H., W.H. Sproat, B.D. Dodd, and J.M. Hamilton, “Reliability of Nondestructive Inspections – Final Report. Technical Report SA-ALC/MME 76-6-38-1, San Antonio Air Logistics Center, 1978.

Moore, D.G. and F.W. Spencer, Interlayer crack detection results using sliding probe eddy current procedures, *10th Asia-Pacific Conference on Non-Destructive Testing*, Brisbane, Australia, 2001.

Packman, P.F., H.S. Pearson, G.B. Marchese, and J.S. Owens, “The Application of a Fracture Mechanics – Nondestructive Testing Design Criterion,” Technical Report AFML-TR-68-32, Air Force Materials Laboratory, 1968.

Rummel, W.D., P.H.J. Todd, S. A. Frecska, and R.A. Rathke, “The Detection of Fatigue Cracks by Nondestructive Testing Methods,” Technical Report NASA CR 2369, National Aeronautics and Space Administration Martin Marietta Aerospace, 1974.

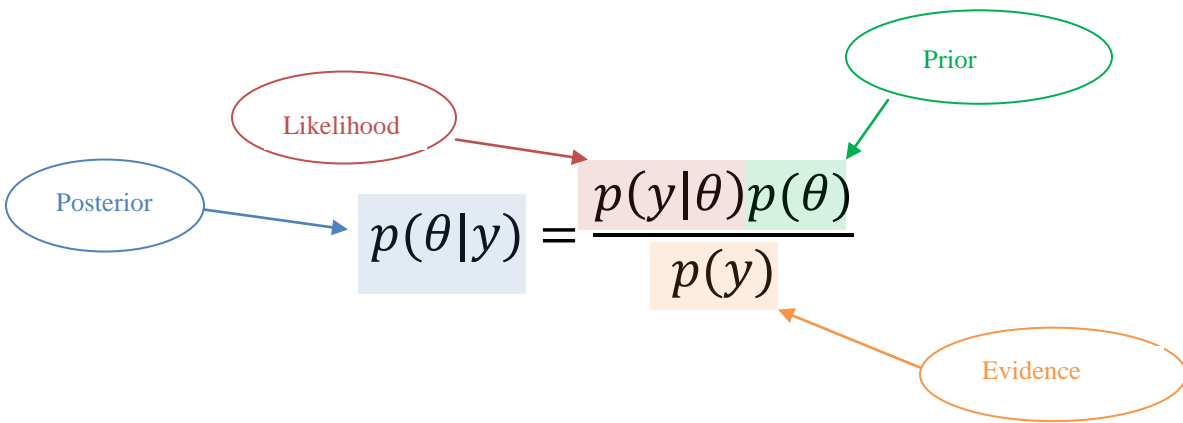
Rummel, W.D., "Recommended Practice for Demonstration of Nondestructive (NDE) Reliability on Aircraft Production Parts," *Materials Evaluation*, Vol. 40, 1982, pp. 922–932.

Rummel, W.D. "NDE Procedure Validation and Use in NDE System Calibration for NDE Applications", *Rev. Prog. Quant. Nondestr. Eval*, Vol 24, 2005, pp. 1982-1986.

Spencer, F.W. "Nonparametric Pod Estimation for Hit/miss Data: a Goodness of Fit Comparison for Parametric Models", *Rev. Prog. Quant. Nondestr. Eval*, Vol 30, 2011, pp. 1557-1564.

Thompson, R.B., A Bayesian Approach to the Inversion of NDE and SHM Data", *Rev. Prog. Quant. Nondestr. Eval*, Vol 29, 2010, pp. 679-686.

U.S. Department of Defense, MIL-HDBK-1823A Nondestructive Evaluation System Reliability Assessment, (August 2010).



- **Prior** – Belief about the statistical parameters or expert opinion.
- **Evidence** – Normalizing constant or marginal likelihood that is important for model selection purposes.
- **Posterior** – Integration of experimental data and beliefs, expert opinion, and physics based models
- **Likelihood** – Probability that the data is generated from statistical model with parameters

Figure 1: Diagram of Bayesian approach used for statistical inference.

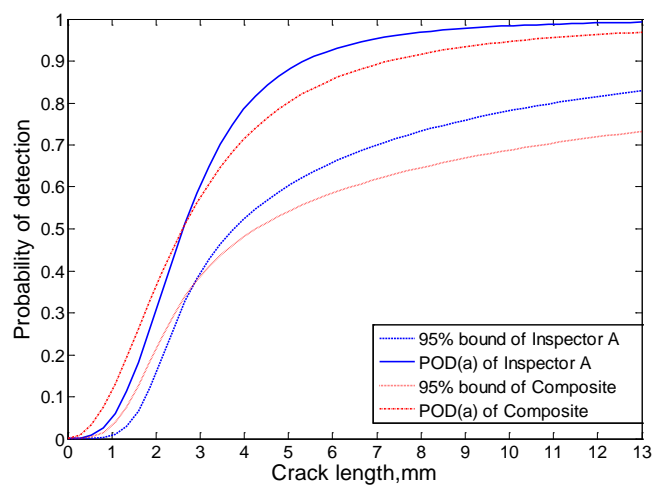


Figure 2. Data from (Berens, 1989) analyzed using Bayesian approach with noninformative priors.

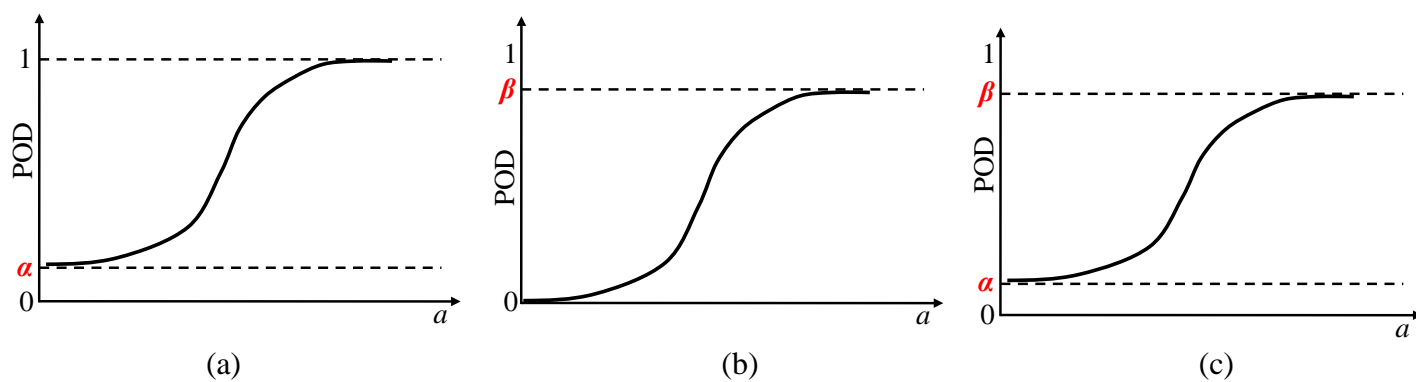


Figure 3. POD model options: (a) 3-parameter with lower bound, (b) 3-parameter with upper bound, and (c) 4-parameter with lower and upper bound.

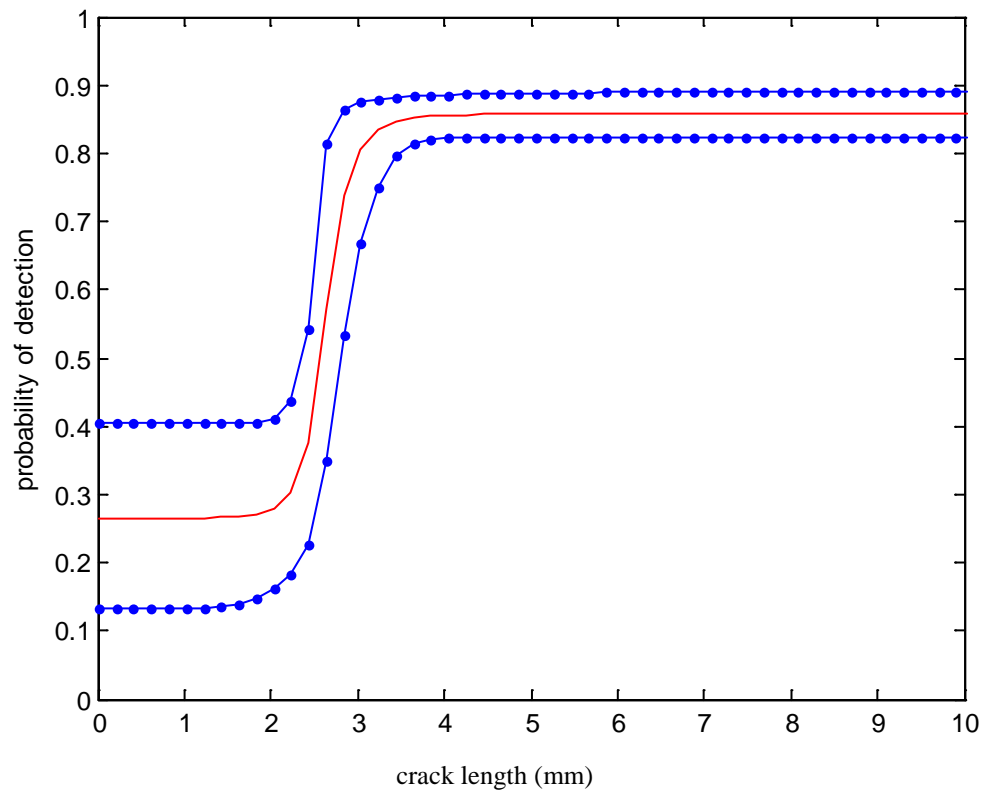


Figure 4. Probit 4-parameter model that best fits data set D8001(3)L.